XmR Control Charts and Data Normality

Forrest Breyfogle
Smarter Solutions, Inc.
www.smartersolutions.com

February 15, 2004

Purpose

The purpose of this paper is to illustrate that XmR control charts are not robust to non-normally distributed data. An alternate control charting approach is presented along with a procedure to describe process capability/performance reporting in terms that are easy to understand and visualize.

Application Example

In the accounts receivable department, invoices are sent to customers for payment. The difference between payment date and due date often follow a log normal distribution. The following data could be considered a random selection of one invoice daily for 1000 days, where payment date for the invoice was subtracted from its due date; e.g., a value of 10 indicates that an invoice payment was 10 days late.

Analysis

1000 points were randomly generated in Minitab from a lognormal distribution with a location parameter of 2, a scale parameter of 1.0, and a threshold of 0. A histogram of the data yielded the plot shown in Figure 1.

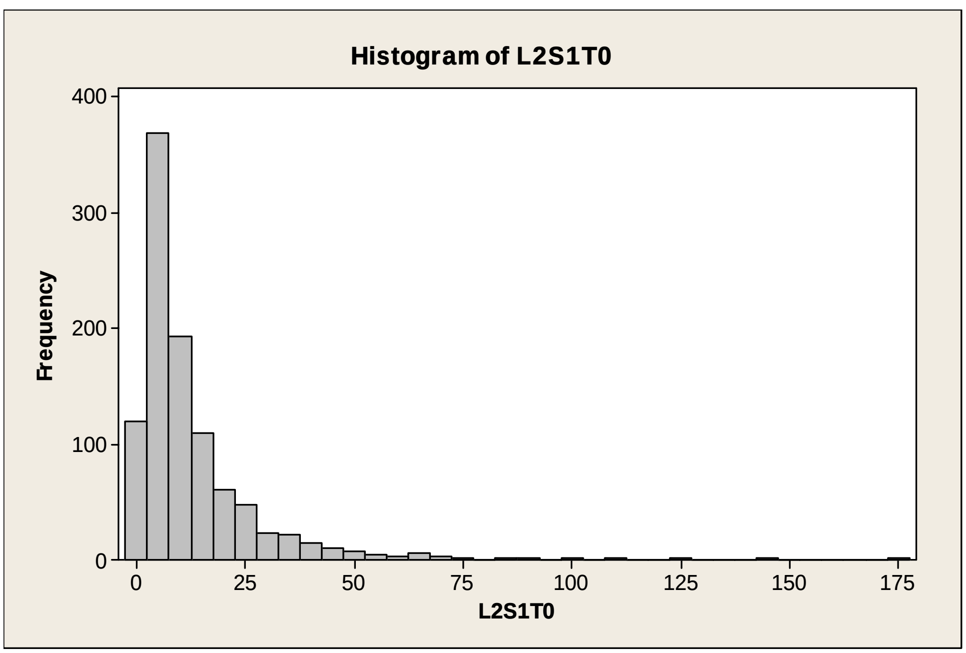

Figure 1: Histogram of 1000 Randomly Generated Points from a Log-Normal Distribution

A normal probability plot of the data is shown in Figure 2

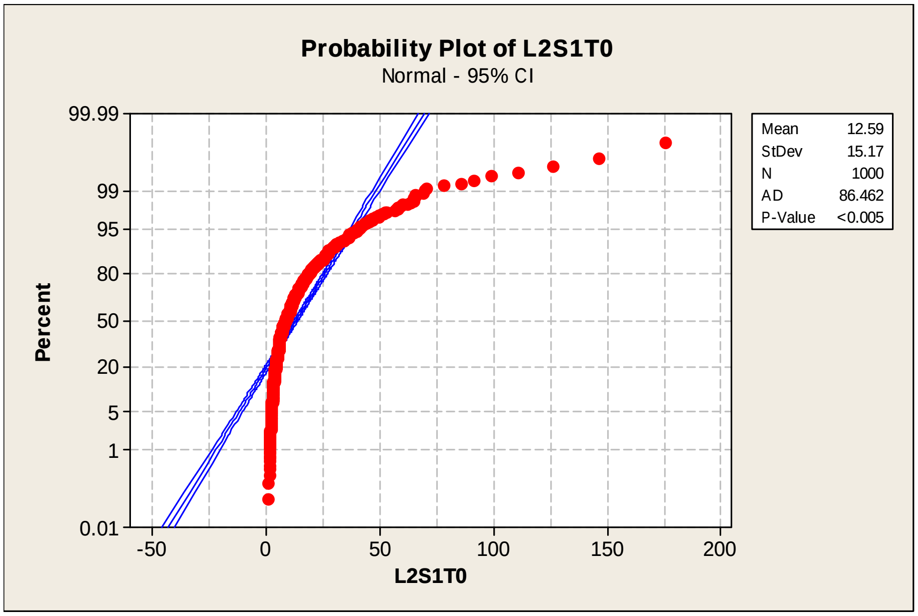

Figure 2: Normal Probability Plot of the Data

From Figure 2, we statistically reject the null hypothesis of normality technically because of the low p-value and physically since the normal probability plotted data does not follow a straight line. This is also logically consistent with the problem setting, where we do not necessarily expect a Normal distribution for the output of such a process. A lognormal probability plot of the data is shown in Figure 3.

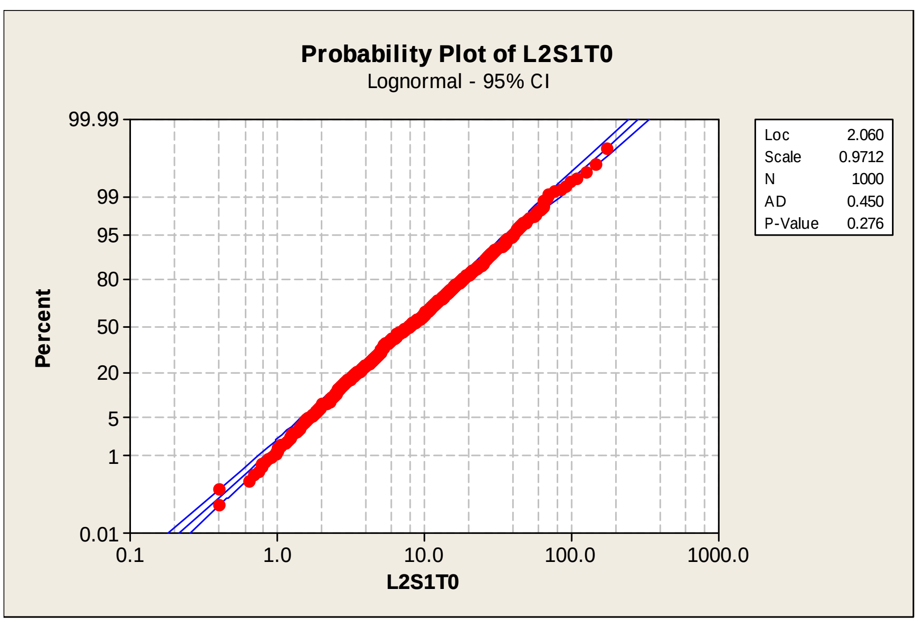

Figure 3: Lognormal Probability Plot of the Data

From Figure 3, we statistically fail to reject the null hypothesis of the data being from a lognormal distribution since the p-value is not below our criteria of 0.05 and physically since the lognormal probability plotted data tends to follow a straight line. Hence, it is reasonable to model the distribution of this variable as Lognormal.

If the XmR control chart (Note: Minitab uses the notation ImR instead of XmR) is robust to the non-normality of data, an XmR control chart of the randomly generated data should be in statistical control. In the most basic sense, using the simplest run rule (a point is "out of control" when it is beyond the control limits) we would expect such data to give a false alarm about 3 or 4 times out of 1000 points. Further, we would expect false alarms below the lower control limit (LCL) to be equally likely to occur as would false alarms above the upper control limit (UCL).

Figure 4 shows an XmR plot of the randomly generated data.

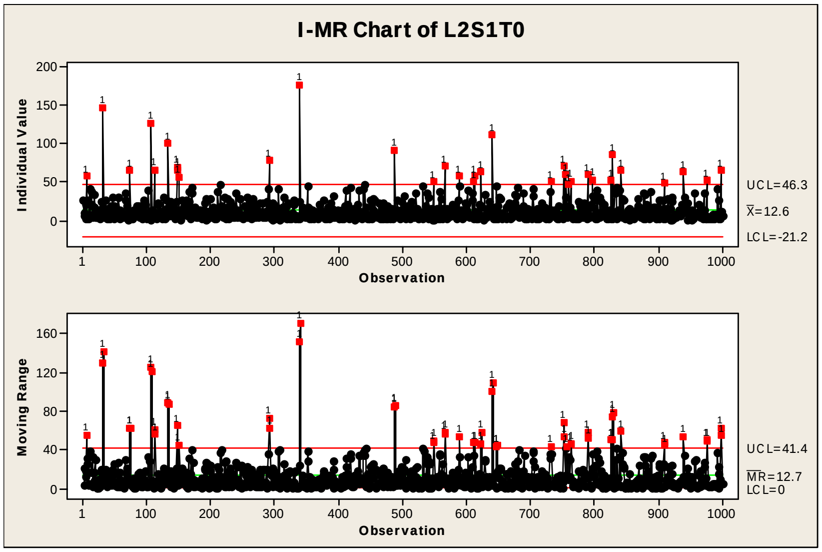

Figure 4. XmR control chart of the data

The individuals control chart in Figure 4 shows many out of control points beyond the upper control limit (UCL). In addition, the individuals control chart shows a physical lower boundary of 0 for the data, which is well within the lower control limit (LCL) of -21.2. We would like to see a random scatter pattern within the control limits, which is not prevalent in the individuals control chart.

Figure 5 shows a control chart using a Box-Cox transformation with a lambda value of 0, the appropriate transformation for log-normally distributed data.

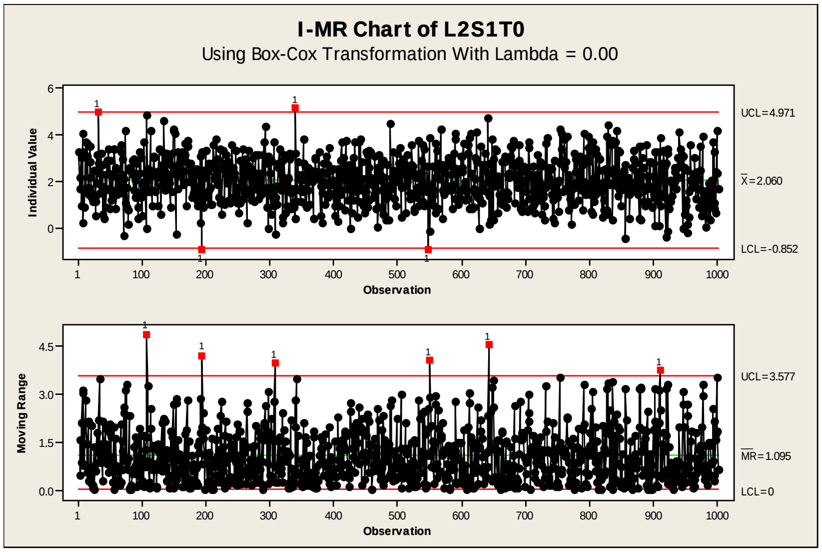

Figure 5 XmR Control Chart with a Box-Cox transformation lambda value of 0

The control chart in Figure 5 is much better behaved than the control chart in Figure 4. Almost all 1000 points in the individuals control chart of the XmR plot are in statistical control. The number of false alarms is consistent with the design and definition of the XmR control chart control limits.

A very descriptive estimate for the process capability/performance metric for this process, when no specification exists, is shown in Figure 6.

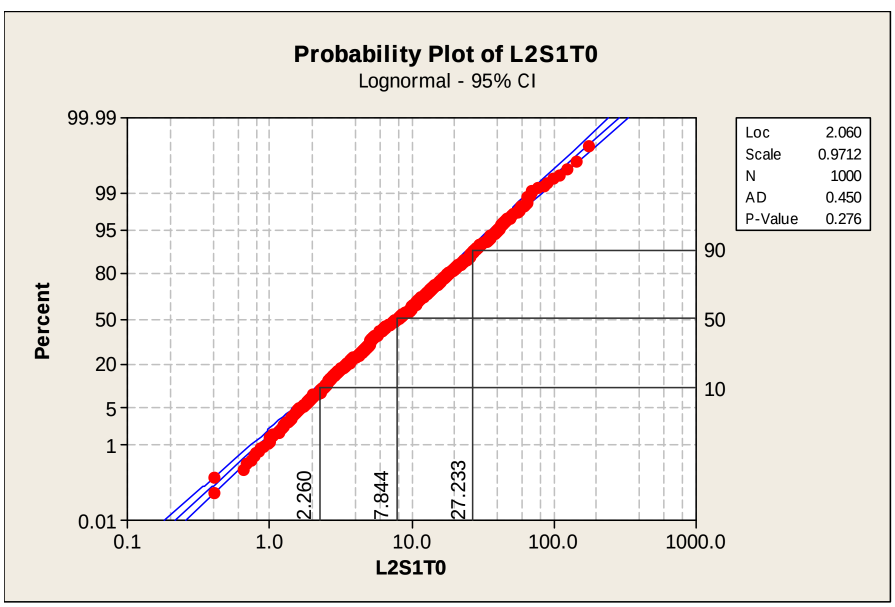

Figure 6: Lognormal plot of data with 80% frequency of occurrence rate

From this Figure 6, a best estimate process capability/performance metric output for the process is 80% of all invoices are paid between 2.3 and 27.2 days beyond the due date, with a median of 7.8 days.

Conclusions

If data are not from a normal distribution, an XmR control chart can generate many false signals, leading to unnecessary tampering with the process. When no specifications exist, a best estimate for the 80% frequency of occurrence rate along with median response is an easy-to-understand description that conveys what the process is expected produce in terms that everyone can visualize.

Reference

Breyfogle, F. W. (2003), *Implementing Six Sigma: Smarter Solutions using Statistical Methods*, 2nd edition, Wiley, New York, NY.

Deming, W. E. (1986), *Out of the Chris*, MIT Press, Cambridge, MA.

Montgomery, D. C. (2000), *Introduction to Statistical Quality Control*, 4th edition, Wiley, New York, NY.